An Empirical Model Describing the Evolution of Cancerous Cells

Malcolm Callis; Colin Brooks; David Liebner, MD[1,2]
[1]Department of Biomedical Informatics, [2]Department of Internal Medicine, The Ohio State University

THE OHIO STATE UNIVERSITY
COLLEGE OF MEDICINE

Abstract

The purpose of this study was to produce a computational model that can simulate the evolution of cancer. It is well known that cancerous cells result from accumulated mutations in genes that cause the cells to divide much more frequently than ordinary cells. Previous simulations that this model is based on have used hypothetical sets of genes. What sets the model presented in this study apart is that it uses genomes from melanoma cases to predict how much a cell with a certain set of mutations will divide. Since the model is based on real world data instead of theoretical genes it should produce a much more accurate model. Hopefully, this model will be able to provide a better understanding of how cancers develop over time and will help determine which genes are significant in various types of cancer. This model can potentially be used to determine which genes to target with drugs.

Introduction

Recent work in bioinformatics has allowed readings of entire genomes of tumors (1). The model presented in this study uses genomes from 346 real world melanoma cases provided by the TCGA database. Most mutations are passenger mutations which do not affect cell growth (1, 2, 3). From the over 18,000 mutated genes from the TCGA database, MutSigCV was used to find the most significantly mutated genes based on false discovery rates to use in the model (2).

The Model

Every generation the model follows the four steps shown in Figure 1.

1. Cell Strains

The entire tumor population is divided into individual cell strains. Each cell strain is composed of cells with identical genetic information. Each cell strain has a population, a fitness and a list of all genes that are mutated in it. The fitness indicates the ratio of cell births to cell deaths.

2. Calculate Population

The strain population for the next generation is equal to the previous population multiplied by the fitness.

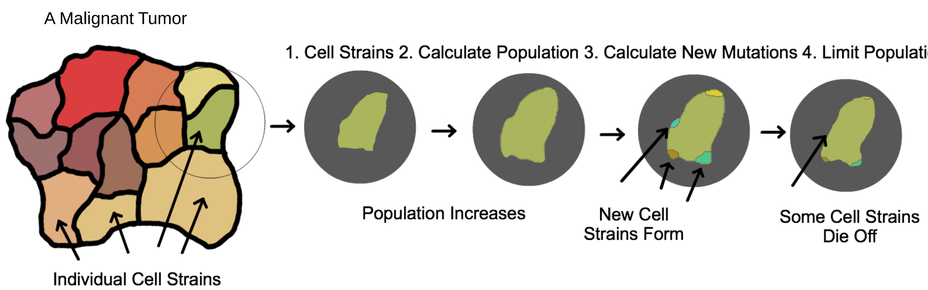

Figure 1. Overview of the Algorithm

The Model (cont.)

3. Calculate New Mutations

Every new cell has a probability of gaining a mutation. When this happens a new cell strain must be created that will subsequently have a new fitness calculated. It is assumed that simulated strains that are similar to strains from the reference set will have higher fitness. First, the strain is compared to every strain in the reference set using a distance metric. Then these distances are used to calculate the kernel density estimation. This estimation is then compared to the expected kernel density estimation given a reference data set generated totally randomly. Finally, this ratio is used to calculate the fitness for the new strain.

4. Limit Population

The model simulates a cubic centimeter of cells which equates to 10^8 cells. If the total population goes above this level the population of each cell strain will be reduced uniformly until the total population reaches 10^8. This creates competition between cell strains in which strains with higher fitness will proliferate and strains with lower fitness will slowly die off, simulating cancer evolution.

Results

Table 1 shows the gene interactions between the ten most significantly mutated genes. For example, the cell at the intersection of BRAF and C15orf23 has a value of 1.768, which indicates that a strain with mutations in these two genes matches the reference set 1.768 times better than a randomly generated set.

Table 1. Ratio of Observed Kernel Density Estimation/Expected Kernel Density Estimation for Different Gene Combinations

Second Mutation \ First Mutation	BRAF	NRAS	TP53	CDKN2A	C15orf23	STK19	OXA1L	ANKRD20A4	RPS27	PTEN
BRAF	1.835									
NRAS	1.893	1.122								
TP53	1.819	1.177	1.025							
CDKN2A	1.855	1.188	1.094	1.047						
C15orf23	1.768	1.117	1.024	1.051	0.974					
STK19	1.897	1.16	1.069	1.091	1.022	1.022				
OXA1L	1.776	1.144	1.056	1.077	1.011	1.054	1.01			
ANKRD20A	1.633	1.042	0.958	0.977	0.915	0.955	0.946	0.904		
RPS27	1.815	1.145	1.052	1.074	1.006	1.05	1.038	0.939	1.002	
PTEN	1.801	1.146	1.045	1.073	1.001	1.043	1.032	0.918	1.027	0.999

Conclusion and Plans for Future Work

This model provides a novel way to simulate cancer which tracks individual genes in order to better understand their evolution. Some limitations of this model are that it treats all mutations in the same gene as identical. In reality, the BRAF and NRAS genes are involved in similar gene pathways and almost never co-mutate. However, this model currently predicts that a strain with a BRAF and an NRAS mutation would have relatively high fitness. However, later iterations of this project plan to use network smoothing to account for this. In addition, later models are planned to simulate the cells in a 3D lattice to allow strains to diverge and to show the way cancer spreads.

References

1. Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K., Vogelstein, B. and Nowak, M. (2010). Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences*, 107(43), pp.18545--18550.
2. Lawrence, M. et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 214-218.
3. The results shown here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.